

소형 스크린 장치를 위한 XML 문서 변환

이영호[○] 김상훈

세명대학교 일반대학원 전산정보학과[○], 세명대학교 소프트웨어학과
{lyh[○], ksh}@plac.semyung.ac.kr

XML document transformation for small-screen device

Young-Ho Lee[○] Sang-Hoon Kim

Dept. of Computer & Information, Semyung University[○], Dept. of Software, Semyung University

요약

PDA나 휴대폰 같은 모바일 장치는 데스크 탑 컴퓨터에 비해 상대적으로 작은 스크린을 가진다. 웹 문서는 일반적으로 데스크 탑 컴퓨터 스크린을 기준으로 작성되기 때문에 이러한 웹 문서가 모바일 장치에서 표현되기 위해서는 다수의 좌, 우 스크롤을 필요하게 된다. 이는 문서 내용의 독이성을 떨어뜨리며 사용자에게 많은 불편함을 준다. 본 논문에서는 XML로 작성된 웹 문서를 소형 스크린 장치에 적합하도록 변환하는 주석 기반 XML 변환 시스템을 제안한다. XML 변환 시스템은 분할 정보가 기술된 외부 주석을 이용하여 XML 문서를 소형 스크린에 알맞은 다수의 XML 문서로 분할하고, 분할된 XML 문서를 HTML 문서로 변환하는 과정으로 XML 문서 변환을 처리한다.

1. 서론

PDA나 휴대폰 같은 모바일 장치에서 인터넷 접속이 가능해짐에 따라 기존의 웹 문서들을 이들 장치에 맞게 변환하는 것이 필요하게 되었다. 기존의 웹 문서들은 데스크 탑 컴퓨터에서 표현되도록 제작되었기 때문에 소형 스크린을 가진 모바일 장치에 맞게 웹 문서들을 변환시키지 않는다면 모바일 장치 사용자는 많은 조작을 해야 될 것이다. 소형 스크린 장치들을 위한 웹 문서 변환 방법으로서 다음과 같이 두 가지 경우가 있다. 첫 번째는 명시적으로 제공된 메타 정보를 이용하여 변환하는 것이고, 두 번째는 메타 정보 없이 자동 변환하는 것이다. 메타 정보인 주석은 변환 대상이 되는 문서에 삽입되거나 원본 문서와 분리된 파일에 존재할 수 있다. 메타 정보를 이용한 변환 시스템으로서 주석 기반 트랜스코딩 시스템(Annotation-Based Transcoding System)은 페이지 분할 힌트를 제공하는 외부 주석을 이용하여 하나의 HTML 문서를 여러 HTML 문서로 분할하였다[1]. 반면 메타 정보 없이 변환하는 시스템으로서 Digstor 시스템은 경험적 방법을 사용하여 HTML 문서를 자동으로 제작하였다[2].

소형 스크린 장치들을 위한 웹 문서 변환의 기존 연구는 대부분 HTML 문서를 대상으로 하였다. HTML은 웹 문서 제작 언어로서 지금까지 많이 사용되어 왔지만 문서의 표현 부분과 구조 부분이 완전히 독립적이지 않으며 사용자 정의 요소를 만들 수 없다는 단점들로 인해 XML이 HTML을 빠르게 대체하고 있는 추세이다. XML은 SGML의 서브 셋으로서 SGML과 HTML 두 개 모두와 상호 운영이 가능하도록 설계되었다. XML 문서는 만들기가 쉬우며 HTML 문서와 동일한 방법으로 웹에서 교환 및 처리될 수 있다[3]. 본 논문에서는 소형 스크린을 가진 모바일 장치를 위해 XML 문서를 변환하는 방법으로서 주석을 이용하여 기존의 XML 문서를 여러 XML 문서들로 분할하

는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서 XML 변환 시스템의 전체 구조와 XML 문서 분할 방법을 설명한다. 3장에서는 제안 시스템의 구현을 통해 변환된 XML 문서들을 실제로 PDA에 적용시킨 결과를 평가하였다. 마지막으로 4장은 결론 및 향후 연구 과제로 이루어져 있다.

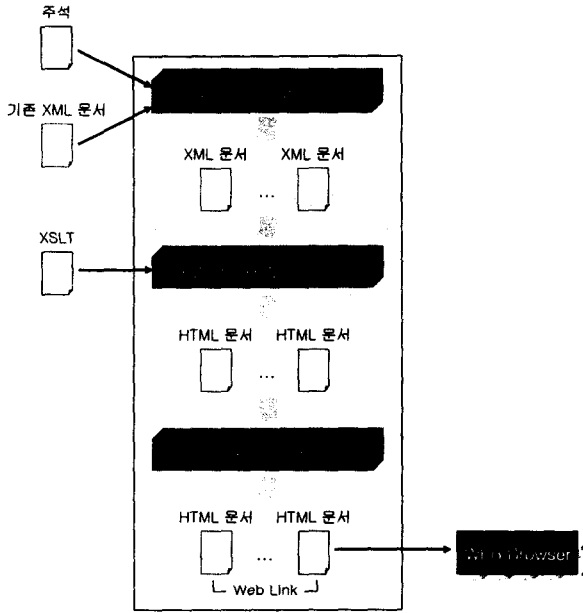
2. 주석 기반 XML 변환 시스템의 설계

많은 양의 자료를 가진 XML 문서가 소형 스크린 장치에서 표현된다면 사용자는 XML 문서를 보기 위해 다수의 좌, 우 스크롤을 해야 될 것이다. 그러므로 본 논문에서는 화면에 표현되는 자료의 양을 줄이기 위해 XML 문서를 분할하는 주석 기반 XML 변환 시스템을 설계하였다.

2.1. XML 변환 시스템의 구조

주석 기반 XML 변환 시스템은 XML 문서 분할기, XML_to_HTML 변환기 그리고 HTML 문서 웹 링크 작성기로 구성되어 있다. XML 문서 분할기는 분할 정보를 제공하는 메타 정보인 외부 주석을 이용하여 하나의 XML 문서를 여러 XML 문서들로 분할한다. 분할된 XML 문서들은 XML_to_HTML 변환기에서 XSLT를 이용하여 HTML 문서들로 변환된다. XSLT는 분할된 XML 문서마다 따로 존재하는 것이 아니라 하나의 XSLT를 이용한다. 만약 원본 XML 문서에 대한 XSLT가 존재한다면 그 XSLT를 이용할 수 있다. 원본 XML 문서를 HTML 문서로 변환하기 위해 제작된 XSLT는 분할에 대한 고려 없이 제작되었을 것이다. 그러므로 XSLT에 의해 만들어지는 웹 링크들 중에서 동일 문서 내를 이동하기 위해 만들어진 웹 링크들은 분할이 완료된 시점에서 목적지 앵커가 다른 문서로 변경 될 수 있기 때문에 이러한 목적지 앵커에 대한 URI를 수정하는 것이 필요하다. 이와 같은 수정에 대한 처리와 함께 변환된 HTML 문서들을 연결하기 위한 새로운

웹 링크들을 만드는 것이 HTML 문서 웹 링크 작성기이다. [그림 1]은 XML 변환 시스템의 전체 구조이다.

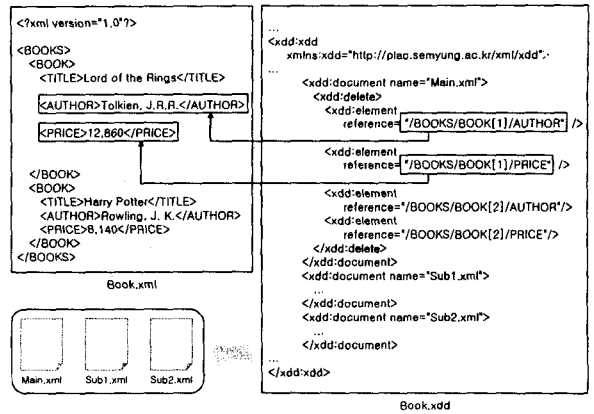


[그림 1] XML 변환 시스템 구조

2.2. XML 문서 분할 방법

XML 변환 시스템은 XML 문서로 작성된 외부 주석을 이용하여 XML 문서를 분할한다. 본 논문에서는 외부 주석으로서 XDD(XML Divider Description)를 사용하였다. XDD는 본 논문에서 XML 문서 분할을 위해 설계되었으며 미리 정의된 요소들을 사용하여 작성한다. XDD 요소의 태그 이름은 xdd:element와 같이 제한된 이름을 사용하며 네임스페이스 접두사는 xdd이다.

XDD는 XML 문서를 분할하기 위해 삽입, 삭제, 갱신의 3가지 연산과 7가지의 연산 대상을 가진다. 3가지 연산과 7가지 연산 대상은 모두 XDD에서 미리 정의된 요소로 지시된다. 3가지 연산을 지시하는 XDD의 요소는 insert, delete, update 요소이다. 삽입 연산을 지시하는 insert 요소의 의미는 분할 대상인 XML 문서에 XDD에서 정의된 새로운 자료를 삽입하라는 것이며 delete 요소와 update 요소의 의미는 각각 XML 문서에 있는 기존 자료를 삭제하고 변경하라는 것이다. 7가지의 연산 대상을 지시하는 XDD의 요소는 pi, element, comment, tag, content, attribute, namespace 요소로 연산을 지시하는 요소의 자식 노드로 나온다. 연산 대상을 지시하는 요소들은 XPath 또는 XPointer를 사용하여 분할 대상인 XML 문서의 일부를 참조하며 XML 문서로 삽입되거나 XML 문서에서 삭제, 변경 되는 자료들에 대한 정보를 제공한다. 예를 들어 삭제 연산에 대한 연산 대상이 content 요소로 지시되었고 XPath에 의해 XML 문서의 어떤 요소 노드가 참조되었다면 그 요소의 모든 내용을 삭제하라는 의미가 된다. [그림 2]는 XML 문서를 분할하기 위해 작성된 XDD 예이다.



[그림 2] XDD 작성 예

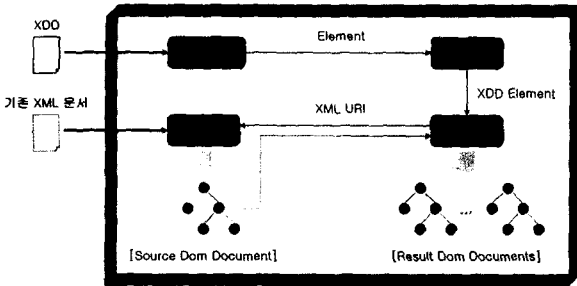
3. XML 변환 시스템의 구현 및 평가

XML 변환 시스템은 리눅스 운영체제에서 J2SE 1.5.0 Beta 1을 이용하여 구현하였다. 그리고 구현된 XML 변환 시스템의 변환 결과를 평가해보기 위하여 Pocket PC 2003 Emulator를 사용하였다.

XML 변환 시스템은 XML 문서를 분할하기 위해 작성된 XDD와 XML 문서를 HTML 문서로 변환하기 위해 작성된 XSLT를 입력받는다. 입력받은 XDD는 XML 문서 분할기로 전달되어 SAX 파서에 의해 처리된다. SAX 파서는 XDD를 처리하면서 특정 이벤트가 발생하면 그 이벤트를 처리하기 위한 핸들러 클래스의 함수들을 호출한다. 핸들러 클래스는 SAX 파서로부터 XDD의 요소가 시작되거나 끝날 때 마다 그 요소를 전달받는다. SAX 파서로부터 전달받은 XDD 요소들은 미리 정의된 의미대로 올바르게 작성되었는지에 대한 검사가 먼저 이루어진다. 전달받은 요소가 XDD에 정의된 요소이고 올바르게 작성되어있었다면 요소의 의미가 분석된 후에 그 의미대로 처리된다.

변환 대상인 XML 문서는 XDD에서 description 요소의 about 속성으로 지시된다. XML 문서 분할기는 description 요소의 about 속성에서 지시된 XML 문서를 DOM 파서로 처리하여 DOM Document 객체로 만든다. Document 객체는 XDD의 document 요소에 의해 새로운 Document 객체로 만들어진다. 새로운 Document 객체는 document 요소의 자식 요소인 insert, delete, update 요소들에 의해 새로운 자료가 삽입되거나 기존 자료가 변경 또는 삭제되고 난 후 XML 문서로 만들어지게 된다. 결과적으로 document 요소가 나올 때 마다 새로운 XML 문서가 만들어지는 것이며 이렇게 만들어진 XML 문서들이 기존 XML 문서가 분할된 것이다. 원본 XML 문서의 분할 결과로 만들어진 XML 문서들은 하나의 메인 문서와 이를 제외한 서브 문서들로 구분된다. 메인 문서는 사용자 브라우저에 처음으로 보여 지는 문서로서 최종적으로 변환된 HTML 문서에 서브 문서들과의 웹 링크를 정의하는 앵커들이 만들어진다. XML 문서 분할기는 메인 문서를 만든 후에 서브 문서들이 만들어 질 때 마다 메인 문서와 현재

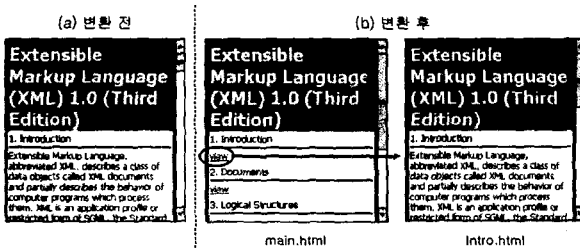
만들어진 서버 문서의 웹 링크 정보를 만든다. 이 웹 링크 정보는 HTML 웹 링크 작성기에서 변환된 HTML 문서들을 연결하기 위한 새로운 웹 링크들을 만드는데 이용된다. [그림 3]은 XML 문서 분할기의 구조이다.



[그림 3] XML 문서 분할기 구조

분할된 XML 문서들은 XML_to_HTML 변환기로 입력된다. 입력된 XML 문서들은 차례대로 DOM 파서에 의해 처리되어 DOM Document 객체로 만들어지게 된다. DOM Document 객체는 처음에 입력받은 XSLT를 이용하여 만들어진 Transformer에 의해 HTML 문서로 만들어진다.

XML_to_HTML 변환기에서 만들어진 HTML 문서들은 HTML 웹 링크 작성기로 입력된다. HTML 웹 링크 작성기는 입력된 HTML 문서들에 존재하는 웹 링크들의 목적지 앵커에 대한 URI를 수정하고 메인 문서가 변환된 HTML 문서와 서버 문서들이 변환된 HTML 문서들과의 웹 링크를 정의한다. 웹 링크들의 목적지 앵커에 대한 URI를 수정하기 위하여 HTML 문서의 모든 A 요소가 검사된다. A 요소의 href 속성 값인 URI가 변환되기 전의 HTML 문서를 지정하고 있다면 URI 끝에 #으로 시작하는 앵커 이름을 추출한다. 추출된 앵커 이름을 가진 앵커를 찾기 위해 모든 HTML 문서들이 차례대로 검색된다. 앵커를 찾았다면 A 요소의 href 속성 값은 앵커가 있는 HTML 문서에 대한 URI로 수정된다. 존재하는 웹 링크들에 대한 URI 수정 처리가 모두 끝났다면 HTML 웹 링크 작성기는 XML 문서 분할기에 의해 제공된 메인 문서와 서버 문서들의 웹 링크 정보를 이용하여 메인 문서가 변환된 HTML 문서에 새로운 앵커들을 만든다. 새로 만들어진 앵커들의 목적지 앵커는 서버 문서들이 변환된 각각의 HTML 문서가 된다.



[그림 4] XML 변환 결과

[그림 4]는 Pocket PC 2003 Emulator에서의 XML 변환 결과를 보여준다. [그림 4]에서 왼쪽 그림은 XML 문서 변환 없이 브라우저에 표현한 것이고 오른쪽 그림은 변환하여 브라우저에 표현한 것이다. main.html이 표현된 것을 보면 변환되기 전의 원본 문서 내용 중에서 각 제목에 해당하는 본문이 삭제되고 그 자리에 새로운 웹 링크가 만들어진 것을 알 수 있다. main.html에서 새로 만들어진 웹 링크들의 목적지 앵커는 삭제된 본문이 저장된 HTML 문서이다. [그림 4]에서 보여주는 것과 같이 변환 후의 HTML 문서들은 원본 HTML 문서에 비해 저장되는 자료의 양이 적기 때문에 전체 문서 길이가 줄어든다.

4. 결론 및 향후 연구 과제

현재 존재하는 대부분의 웹 문서는 데스크 탑 컴퓨터에서 표현되도록 제작되었다. 그러므로 소형 스크린을 가진 모바일 장치에서 이러한 웹 문서를 보는 것은 장치 사용자의 많은 조작을 요구하게 된다. 본 논문에서는 소형 스크린을 가진 모바일 장치를 위해 외부 주석을 이용하여 XML 문서를 분할하는 주석 기반 XML 변환 시스템을 설계하고 구현하였다. XML 변환 시스템은 기존 XML 문서를 분할하는 XML 문서 분할기, 분할된 XML 문서들을 HTML 문서들로 변환하는 XML_to_HTML 변환기, 그리고 HTML 문서들의 웹 링크들을 수정하거나 새로 만드는 HTML 웹 링크 작성기로 구성되었다.

XML 변환 시스템은 XML 문서를 분할하기 위하여 외부 주석을 이용한다. 현재 정의된 외부 주석은 분할 정보를 명확하게 기술할 수 있다는 장점이 있지만 기술해야 할 정보가 많은 경우 주석을 작성하는데 많은 시간이 걸린다는 단점이 있다. 이러한 단점을 극복하기 위한 부분을 향후 연구 과제로 하여 보완할 것이다.

5. 참고문헌

- [1] M. Hori, etc, Annotation-Based Web Content Transcoding, The 9th WWW conference, 2000
- [2] T. W. Bickmore, B. N. Schilit, Digester: Device-independent Access to the World Wide Web, The 6th WWW conference, 1997
- [3] W3C Recommendation, Extensible Markup Language (XML) 1.0 (Second Edition), 2000
- [4] W3C Recommendation, Namespaces in XML, 1999
- [5] W3C Recommendation, XSL Transformations (XSLT) Version 1.0, 1999
- [6] W3C Recommendation, XML Path Language (XPath) Version 1.0, 1999
- [7] Sun microsystems, Java API for XML Processing Version 1.2, 2002